Insert title here

Samuel John Herniman

Being an

Honours project

submitted to

Bangor University

in partial fulfilment of an

Honours degree in Applied Terrestrial and Marine Ecology

May 2016 500315403

# Declaration

I declare that this is the result of my own investigation and that it has not been submitted or accepted in whole or part for any degree, nor is it being submitted for any other degree.

Candidate: Samuel J Herniman

Signature: ……………………………….

# Acknowledgements

As I think back on where this project started for me, it has been more than just a

# Abstract

The tropical forests of Xishuangbanna are one of the most biodiverse regions of China. Within the last 20 years, 22% of the land has been converted to rubber plantation and tea resulting in large scale fragmentation and habitat loss. In addition, Xishuangbanna’s close proximity to Myanmar, Laos and Vietnam means that it is a hub for wildlife trade, threatening many species in the region.

With over 400 orchid species of 115 genera, Xishuangbanna holds 31% of all Chinese orchids. Orchids are habitat specialists due to their specific pollinators and mycorrhizal associations, and are often vulnerable to many anthropogenic activities. Of these 400, 3 have been classified as possibly extinct in the wild, 15 Critically Endangered, 82 as Endangered and 124 as Vulnerable, regionally.

In this study we examine the distribution of two species that are thought to be threatened by both habitat loss and trade for horticultural use: *Luisia magniflora* and *Dendrobium thyrsiflorum*. Both species are classified as Endangered in a regional Red listing assessment yet little is known about their habitat requirements or distribution. Using existing presence data from digital herbarium records and ecological surveys in Xishuangbanna over the last five years, we produced habitat suitability models of these species using MAXENT. Environmental predictors such as distance to roads, water and settlements, as well as altitude, land cover, canopy cover, aspect, precipitation and daytime temperature were taken into account.

Our results show areas of high probability of presence and we identify environmental factors influencing the distribution of the species. We also highlight the areas of high conservation concern by modeling the future distribution of tea and rubber plantations with MAXENT. By overlaying the existing protected area network onto the distribution maps we can assess the efficacy of protected area coverage.

# Contents

[Declaration 1](#_Toc447649978)

[Acknowledgements 2](#_Toc447649979)

[Abstract 3](#_Toc447649980)

[Contents 5](#_Toc447649981)

[Introduction 7](#_Toc447649982)

[Orchidaceae 7](#_Toc447649983)

[Diversity of Orchids 7](#_Toc447649984)

[Use of orchids globally 7](#_Toc447649985)

[Orchids in Trade 7](#_Toc447649986)

[Sources of orchids 7](#_Toc447649987)

[Demand for orchids 7](#_Toc447649988)

[Horticulture or medicine? 8](#_Toc447649989)

[CITES 8](#_Toc447649990)

[Orchids in China 9](#_Toc447649991)

[Diversity 9](#_Toc447649992)

[Cultural significance 9](#_Toc447649993)

[Use in China 9](#_Toc447649994)

[Horticulture 9](#_Toc447649995)

[Medicine 9](#_Toc447649996)

[Threats to orchids in China 9](#_Toc447649997)

[Wild harvesting 9](#_Toc447649998)

[Deforestation 10](#_Toc447649999)

[Cultivation 10](#_Toc447650000)

[Protected areas 10](#_Toc447650001)

[Description of the study site 10](#_Toc447650002)

[Maximum Entropy 11](#_Toc447650003)

[What is Maximum Entropy? 11](#_Toc447650004)

[What is MAXENT? 13](#_Toc447650005)

[What can it be used for? 17](#_Toc447650006)

[Where it falls short… 18](#_Toc447650007)

[Conclusions 20](#_Toc447650008)

[Methods 21](#_Toc447650009)

[Results 22](#_Toc447650010)

[Discussion 23](#_Toc447650011)

[Conclusions 24](#_Toc447650012)

[References 25](#_Toc447650013)

[Appendices 26](#_Toc447650014)

# Introduction

## Orchidaceae

### Diversity of Orchids

The Orchidaceae family is one of the largest families of flowering plants (sharing the top spot with the Asteraceae), composed of more than 25,000 species and 736 genera (Liu, Luo & Liu 2010; Chase *et al.* 2015).

### Use of orchids globally

Horticulture

Medicine

*Gastrodia elata, Dendrobium offcinale, Ludisia discolor*

Vanilla

## Orchids in Trade

The global orchid trade is vast and varied (Goh & Kavaljian 1989)

### Sources of orchids

### Demand for orchids

Goh & Kavaljian (1989) outlined the factors determining the demand for orchids. These include: whether or not members of a country **consume** flowers regularly, consumer income, energy cost in production of the orchid, vaselife and quality, fashion, and predictions of the importers and distributors (Goh & Kavaljian 1989).

### Horticulture or medicine?

### CITES

The primary mechanism in place to reduce wildlife trade is the Convention on International Trade in Endangered Species of Wild Fauna and Flora (CITES). CITES has been successful in the case of intense farming and regulatory effort in combinations with the SARS epidemic of 2003 (Jiang *et al.* 2013), but its effectiveness has been called into question (Phelps *et al.* 2010; Challender & MacMillan 2014; Nijman & Shepherd 2015). Trade bans can, in some cases, exacerbate overexploitation threats (Conrad 2012) and drive wildlife trade underground rather than reducing it (Veríssimo, Diogo; Challender, Daniel W.S. & Nijman 2012; Biggs *et al.* 2013).

While CITES makes it illegal to trade orchids across borders, it is not illegal to harvest orchids and trade them within China (citation needed). There is a need for research to determine whether the orchids bought and sold around the borders of China were harvested within China or one of its neighbors.

**Where do orchids fall under CITES?**

CITES lists all of the known wild Orchids (Tian, Chen & Xing 2013)

## Orchids in China

### Diversity

There are more than 1200 native orchid species and 173 genera of orchids in China (Liu, Luo & Liu 2010)

### Cultural significance

Orchids have been highly admired in the Chinese culture since the time of Confucius (Goh & Kavaljian 1989).

### Use in China

Orchids are used horticulturally and medicinally in China

### Horticulture

### Medicine

Examples: *Dendrobium officinale* had a large distribution throughout the south of China but is now rare in the wild (Ding *et al.* 2008)

## Threats to orchids in China

The conservation of orchids is threatened by overharvesting and habitat destruction (Liu, Luo & Liu 2010). There is an urgent need for protection (Liu, Luo & Liu 2010).

### Wild harvesting

Example: *Gastrodia elata.* 1960s. Collapse of wild populations due to wild harvesting.

### Deforestation

### Cultivation

Very few native orchids in China have been cultivated on a large scale (Liu, Luo & Liu 2010). This is partly due to the mycorrhizal fungal associations many orchids rely on (Liu, Luo & Liu 2010) (citation needed). Even if cultivation were possible, Williams, Jones & Annewandter (2014) found that some cultivation programmes can increase the strain of wild harvesting. However there are successful examples of this technique (Liu, Luo & Liu 2010).

### Protected areas

Chinese authorities **struggle** to effectively manage protected areas (Zhou & Grumbine 2011) as little enforcement is carried out in these areas and they are largely underfunded.

While the presence of ineffective protected areas often causes people to feel that conservation progress is being made, the presence of a protected are superior to no protected area at all (Zhou & Grumbine 2011)

## Description of the study site

Xishuangbanna Dai Autonomous Prefecture is a area of southern Yunnan, China (21°09’-22°33’N, 99°58’-101°50’E) (Gao: 21°09’-22°36’N, 99°58’-101°50’E) taking up 19,220 square kilometres and supports roughly 12 percent of China’s flora (Shou-qing; Cao & Zhang 1997). Xishuangbanna is neighbors to both Myanmar and Laos (Zhu, Wang & Li 1998).

Xishuangbanna has a tropical monsoon climate, caused by the barrier to northern cold air formed by the Hengduan Mountains in the north of the region (Zhu, Wang & Li 1998; Zhu, Cao & Hu 2006a). In the dry season, lingering fog, caused by the mountainous terrain, supplements the lack of precipitation (Zhu, Wang & Li 1998; Zhu, Cao & Hu 2006a).

Being in the transitional zone between South-East Asia, subtropical East Asia, the Sino-Japanese floristic region, and the Sino-Himalayan floristic region, Xishuangbanna is a biodiversity hotspot (Zhu, Cao & Hu 2006b) (Zhu, Cao & Hu 2006a)

Yunnan happens to be China’s most biologically and culturally rich area (Zhou & Grumbine 2011)

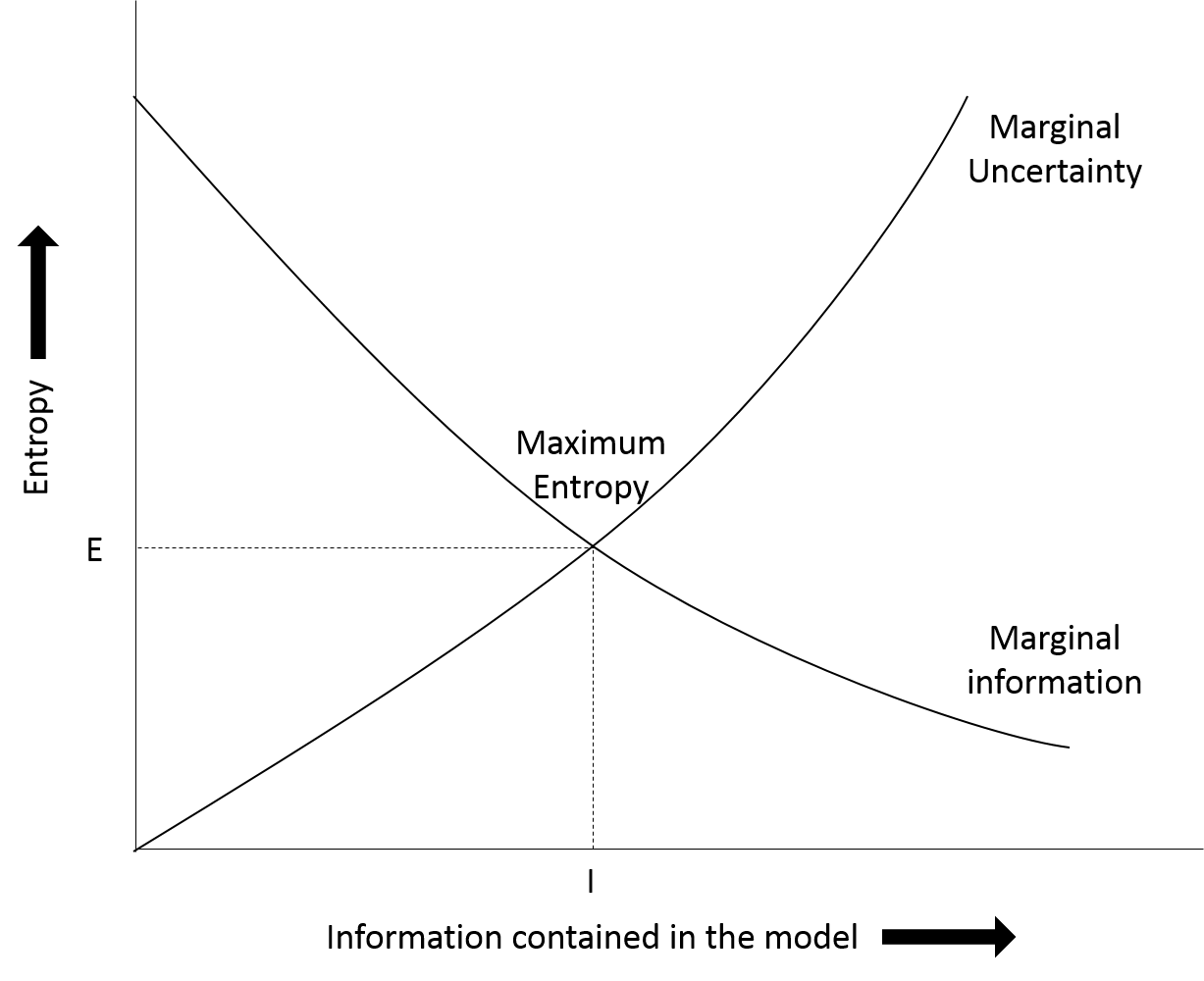
## Maximum Entropy

### What is Maximum Entropy?

To understand MAXENT we must first understand the concept of maximum entropy on which MAXENT is based.

Entropy is a measure of uncertainty or a lack of information (Penfield & Lloyd 2003). ˻When there is no information distinguishing one explanation over another, all the possible explanations can be given equal probabilities and entropy is at its greatest point (log *m* where *m* is the number of possible explanations)˼(Jaynes 1963). This is the principle of indifference (Conrad 2004). At its lowest, entropy has a value of 0, denoting that there is enough information to identify one explanation with no uncertainty over all other possibilities (Jaynes 1963; Penfield & Lloyd 2003; Phillips, Anderson & Schapire 2006).

The principle of maximum entropy is an elementary concept, although it is often explained in a counter-intuitive and convoluted manner: ˻the explanatory distribution (output) must take into account all of the provided data without making any assumptions not backed up by the data˼ (Lahoz Monfort 2008). Essentially, when choosing between a number of explanations for a phenomenon, it is optimal to choose the one that maximises the entropy value while incorporating all the available information about the phenomenon, this is demonstrated in **Figure 1.** (Jaynes 1963; Conrad 2004). Therefore, the maximum entropy distribution is the most informed prediction one can make from the given information and the only prediction that can be made with reason. (Jaynes 1963; Penfield & Lloyd 2003; Conrad 2004).



**Figure 1.** Illustrates entropy vs information contained in the model. Where E is the greatest entropy value possible considering all the available information and I is the point where all available information is incorporated into the model. If a model contains more information than I it is making assumptions and if a model contains less information than I it is ignorantly not using all information available. The point of maximum entropy is where marginal information meets marginal uncertainty.

Thus, MAXENT produces a predicted species distribution over a landscape using all the information available without making any uninformed assumptions. A species is assumed to be present in an area unless there is information given which says otherwise. If no prior presence information was provided to MAXENT, a uniform species distribution would be produced – essentially the null hypothesis (Phillips, Dudík & Schapire 2004; Elith, Kearney & Phillips 2010).

### What is MAXENT?

˻MaxEnt is a presence only, machine learning algorithm used to create habitat suitability maps based on the principle of maximum entropy (Phillips *et al.* 2009). MAXENT associates an input ofgeoreferenced presence points of a certain species with relevant environmental variables which have an influence on the distribution of a species (i.e. temperature, humidity, distance from rivers) to plot their predicted distributions across landscapes (Merow, Smith & Silander 2013; Särkinen, Gonzáles & Knapp 2013). These are referred to as *factors* in MAXENT (Lahoz Monfort 2008). ˻When doing this, MAXENT combines the known presence points and environmental variables to determine what encompasses the species’ ecological niche, producing a potential distribution of the focus species˼ (Phillips, Dudík & Schapire 2004; Phillips, Anderson & Schapire 2006; Williams 2008; Elith, Kearney & Phillips 2010; Araújo & Peterson 2012).

**Receiver operating characteristic (ROC) analysis**

To tell whether a model produced by MAXENT is able to predict a presence point correctly, MAXENT conducts receiver operating characteristic (ROC) analysis (Williams 2008). A curve is produced with sensitivity (true positives) on the *y* axis and 1-specificity (false negatives) on the *x* axis (Phillips, Anderson & Schapire 2006). The area under the curve (AUC) is measured with a value of 1 denoting no incorrect predictions and a value of 0.5 denoting predictions that are no more accurate than random (Elith *et al.* 2006; Williams 2008).

**Presence Only Data**

Rather than looking at presence-absence data, MAXENT takes an input of presence-only data and converts it to presence-pseudoabsence data where the background areas (where no presence point is available) are considered to have an unknown presence – also known as a pseudoabsence (Phillips, Anderson & Schapire 2006; Phillips *et al.* 2009; Merow, Smith & Silander 2013). This presence-pseudoabsence data can then be used to determine the likelihood that an unknown space is occupied or not.

˻The documentation of accurate presence-absence data is relatively inefficient to carry out in most areas of the world. However, natural history museums and herbaria have a plethora of presence-only data˼ (Elith, Kearney & Phillips 2010; Merow, Smith & Silander 2013). Therefore, the use of presence only algorithms, like MAXENT, is realistic and relevant to conservation (Phillips, Dudík & Schapire 2004). MAXENT has gained a following due to its ability to produce accurate distributions based on very few (less than 100) presence only data points (Phillips, Dudík & Schapire 2004; Pearson *et al.* 2007; Merow, Smith & Silander 2013). This presence-pseudoabsence data is combined with environmental variables to produce the MAXENT model.

**Environmental variables (features)**

Since MAXENT attempts to find patterns in the associations between the presence points and environmental variables, careful selection of which environmental variables should be included in the model is required. It is important to exclude variables which have little or no effect on the distribution of the species in question from a MAXENT model to reduce the likelihood of the algorithm finding associations which do not exist. Additionally it is wise to include all the environmental variables which do have an effect on the distribution to gain a complete understanding of the environmental envelopes the species reside in (Merow, Smith & Silander 2013; Hughes 2015).

It is important to note that when MAXENT is used for projecting the distribution of species into the future or studying environmental influences of a species, the feature classes (environmental variables) must be chosen carefully to create an accurate model (Merow, Smith & Silander 2013). However, when MAXENT is used for looking at the accuracy of presence points, MAXENT can determine the most appropriate features to include in the analysis using a machine learning algorithm and so it is possible to include all reasonable layers without affecting the output (Merow, Smith & Silander 2013). This process of determining which features are relevant or not is known as regularisation.

**Regularisation**

Regularisation is the process by which MAXENT removes the features which are not relevant to the model and maximises the use of the relevant ones (Phillips, Anderson & Schapire 2006; Hastie, Tibshirani & Friedman 2009; Merow, Smith & Silander 2013). The default regularisation settings are adequate for most species, however more accurate regularization may be achieved with more appropriate settings (Elith, Kearney & Phillips 2010; Merow, Smith & Silander 2013). A greater regularization value has the effect of removing more features from the analysis. This will create a simplified model but may lack influential features (Merow, Smith & Silander 2013).

### What can it be used for?

Once MAXENT produces a distribution model, it can be useful in many ecological applications. Among them:

**Assessing the possible spread of invasive species**

A distribution of the potential sites that an invasive species could move into is useful for conservationists who wish to take preventative measures to stop this spread from occurring (Araújo & Peterson 2012). It can also be useful to assess why an invasive species has not moved into a certain area when given enough time. This knowledge may lead researchers to a weakness of the invasive species.

**Climate change outlooks for at risk species**

It is common to use MAXENT and related algorithms to make predictions about how a species will be affected by climate change. MAXENT can be used to produce distributions that show where the species will no longer be able to inhabit once changes occur or it may show possible areas where the species may occupy in the future (Araújo & Peterson 2012). It is important to note that these predictions are only accurate assuming that the species makes no adaptations to climate change and is only capable of survival in the climates it already occupies.

**Discovering new populations of a species**

It is possible for researchers to use the known data about a species to determine where it may occur in addition to currently known populations and then visit these potential sites to confirm presence (Araújo & Peterson 2012).

**Determining areas of conservation concern and the most appropriate reserve boundaries**

Knowledge of where a species occurs when delimiting areas to conserve or reserve boundaries is very important since conservation funding is often limited. Making a reserve too large can be an unneeded financial burden while making a reserve too small can impact the survival of the species of concern (Araújo & Peterson 2012).

**Assessing potential areas for reintroductions and translocations**

MAXENT can determine new areas with suitable habitats for successful introductions. ˻As reintroductions are rarely successful˼, it is helpful to know with some certainty that there is a likelihood of success on a particular project (Araújo & Peterson 2012; Ewen, Soorae & Canessa 2014).

### Where it falls short…

Araújo and Peterson (2012) argue that while a clearly stated concept and purpose of using MAXENT can be informative and useful, it is not uncommon for the concept to be vague or unexplained leading to misinterpretation. It is simple to counteract these issues by carefully assessing the reasons for carrying out the modelling, assumptions, and what the potential results are prior to running MAXENT(Araújo & Peterson 2012).

Merow et al. (2013) claim that researchers are often unfamiliar with MAXENT to an appropriate level and rarely make informed decisions when choosing settings. Researchers generally choose the defaults while more appropriate settings (or even algorithms) are unknown (Merow, Smith & Silander 2013).

**Sampling bias**

Sampling bias is inherent in presence only data. It is very common for samplers to survey areas near roads and that are easily accessible. While sampling bias can be accounted for when the search effort is known, a process called Target Group Sampling (TGS) can be used to roughly account for sampling bias when search effort is unknown. TGS uses presence data from similar species to predict the presence of the target species over areas that have not been sampled (Phillips *et al.* 2009; Merow, Smith & Silander 2013).

**Range**

When setting up data before running MAXENT, it is important to choose the range of the analysis wisely. When MAXENT is run over a larger area than the known range of a species, MAXENT will identify suitable habitats for the species outside of the known range (Phillips *et al.* 2009; Webber *et al.* 2011; Barve *et al.* 2011; Merow, Smith & Silander 2013). Whereas running MAXENT on a smaller area than the known range of the species will not give a representative distribution of the suitable habitat of the species. These possibilities are adequate when carrying out appropriate analysis but it is important to consider the question being asked and the goals of the analysis before setting up MAXENT.

### Conclusions

The machine learning algorithm, MAXENT combines presence-only occurrence data of a particular species with relevant environmental layers (factors) to produce a habitat suitability model of the bioclimatic envelopes the species occupies or is capable of occupying. It is important to understand the limitations and capabilities of MAXENT before setting up the software and running the model and to carefully consider the methods and goals of a particular study. With a vast amount of presence-only species occurrence data available in herbaria and natural history museums, MAXENT is a strong tool in conservation.

Methods

# Results

# Discussion

# Conclusions

# References

# Appendices